Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 1772–1792
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Likelihood ratio tests for equality of shape under varying degrees of orientation invariance

Finbarr Holland, Kingshuk Roy Choudhury*

School of Mathematical Sciences, University College Cork, Ireland

Received 7 November 2006
Available online 9 February 2008

Abstract

We consider a problem from image cytometry where the objective is to describe possible changes in the shape and orientation of cellular nuclei after treatment with a toxin. The shapes of nuclei are represented by individual ellipses. It is argued that the shape comparison problem can be formulated as a generalization of a hypothesis test for the equality of covariance matrices. For many cell types, the test statistic should be invariant with respect to orientations of the cells. For other cell types, the test statistic should be equivariant with respect to orientations of the cells, but invariant with respect to orientations of the images. Likelihood ratio tests (LRTs) are derived under a Wishart model. The likelihood maximization uses a new result about the minimization of the determinant of a sum of matrices under individual rotations. The applicability and limitations of these LRTs are demonstrated by means of simulation experiments. The reference distributions of the test statistics under the null hypothesis are obtained using unrestricted and restricted randomization procedures. Justification for the Wishart model is provided using a residual diagnostic method. The scientific implications of the results are considered.

AMS 2000 subject classifications: 57N25; 47N60; 58J70

Keywords: Shape analysis; Likelihood ratio test; Orientation invariance

1. Introduction

The field of cytometry deals with the measurement of cells. A variety of techniques are available for investigating the morphometric properties of cells [10]. In many investigations

* Corresponding address: School of Mathematical Sciences, University College Cork, Statistics Department, UCC Cork, Ireland.
E-mail address: kingshuk@stat.ucc.ie (K. Roy Choudhury).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2008.01.017

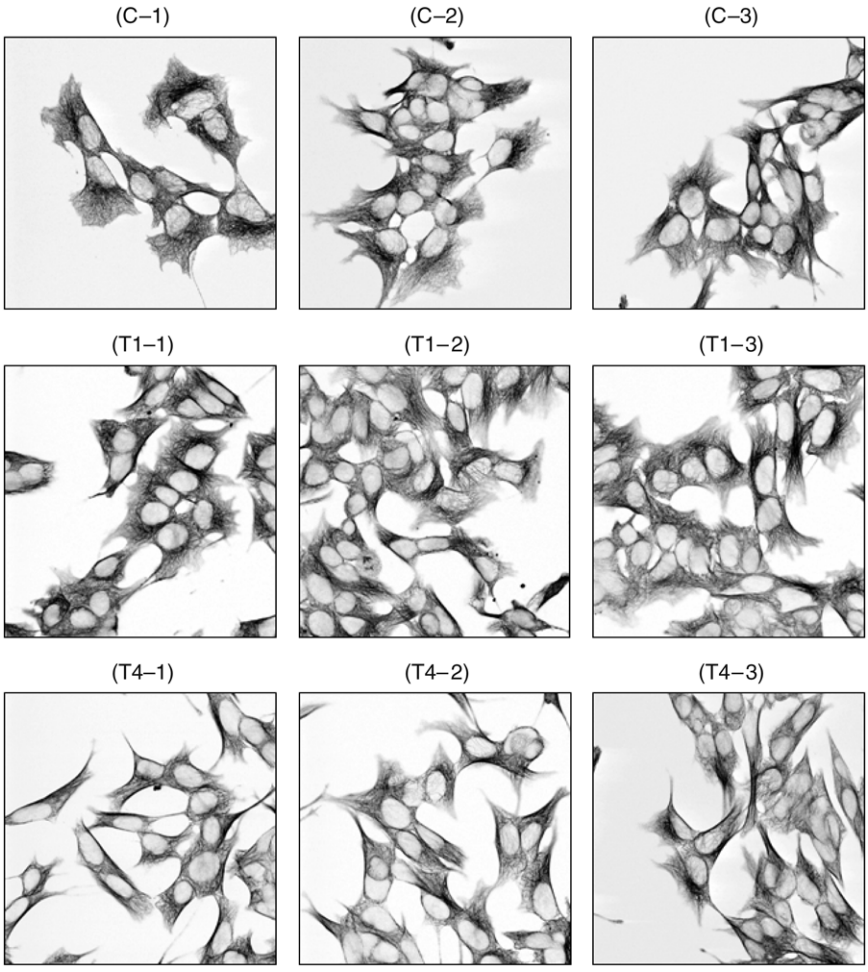

Fig. 1. Images of neuroblast tissue taken using a confocal microscope across different experimental conditions. C-1, C-2 and C-3 are three replicates from the control group (no MPP^+ applied). Images marked T1 are replicates from a group where 5 mM of MPP^+ was applied and the result observed after 1 h. Images marked T4 are replicates from a group where 5 mM of MPP^+ was applied and the result observed after 4 h. Boundaries of nuclei recognized by CASRG are superimposed in red.

involving cytometry, the goal is to describe the effect upon the cells of some change in experimental conditions. An example of such an experiment is a study carried out to investigate the effect of the neurotoxin MPP^+ on the intracellular structure of a human neuroblastoma cell line, SH-SY5Y [8]. The images in Fig. 1 show cross-sections of this culture, stained with β tubulin antibody. The visible organelles are the nuclei (light grey ovoid shapes) and microtubulin structure (darker surrounding material). The goal of the experiment is to characterize changes in cellular structure in response to MPP^+ exposure (at 5 mM levels of concentration) after 1 (T1) and 4 h (T4).

Typically, the data from image cytometry studies consist of measurements on objects $O_{ijk}, i = 1, \ldots, n_{jk}, j = 1, \ldots, I_k, k = 1, \ldots, G$, where the index k runs over experimental

conditions (in the above example $G = 3$). The index j runs over images, or fields of view, per experimental condition (in the above example, $I_k = 3$). Finally, the index i runs over objects within each image. The problem of comparison of shape in cytometry experiments can be formulated as a comparison between the 'average' shape $\mathbf{O}_k$ across experimental conditions. In practice, the method of comparison depends heavily on the way shapes are represented. A number of different representations of object shapes (primarily using the object boundary) have previously been used in the literature, such as: thin plate splines [3], deformation templates [1], polygons [11], or Fourier descriptors [7].

The nuclei in Fig. 1 are mostly ovoid. For such objects, using a fitted ellipse is often an adequate and convenient method of representing their shape. The equation of an ellipse can be written as

$$(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu) = 1. \tag{1.1}$$

Here x is a point in the plane, μ is the centre of the ellipse and Σ is the covariance or 'shape' matrix of the ellipse. Thus, for shape analysis purposes, it is 'sufficient' to consider the 2×2 shape matrices S_{ijk} corresponding to the objects O_{ijk}. In image cytometry, this approach has previously been employed in [10] to study the shapes and sizes of algal cells. Briefly, this involved two steps: (i) identification of the boundaries of each object; (ii) fitting of ellipses by matching moments of the identified regions to those of the ellipse. The boundary identification algorithm used for the objects in Fig. 1 is a variant of seeded region growing, described in [14]. The ellipse fitting is achieved by a method of moments estimation procedure, described in [10]. The form of the estimator will be important, so this description is reproduced here. We consider the object O as a 'point set', i.e. a set of points $p_i = (x_i, y_i)$, $i = 1, 2, \ldots, n$, that lie inside the boundary of O. The (k, l)th moment of O is defined to be $m_{k,l}$:

$$m_{k,l} = \frac{1}{n}\sum_{i=1}^{n} x_i^k y_i^l. \tag{1.2}$$

The first moments $m_{1,0}$ and $m_{0,1}$ jointly describe the centre of location, or *centroid*, $\bar{p}(m_{1,0}, m_{0,1})$ of the object. The second central moments are given by

$$\begin{aligned} m^0_{2,0} &= m_{2,0} - m^2_{1,0}, \\ m^0_{0,2} &= m_{0,2} - m^2_{0,1}, \\ m^0_{1,1} &= m_{1,1} - m_{1,0}m_{0,1}. \end{aligned} \tag{1.3}$$

In these terms, the estimated 2×2 shape matrix S is defined as the covariance matrix

$$S = \begin{pmatrix} m^0_{2,0} & m^0_{1,1} \\ m^0_{1,1} & m^0_{0,2} \end{pmatrix}. \tag{1.4}$$

By and large, the quality of ellipse fit to the objects in Fig. 1 is very good, primarily because the nuclei are essentially elliptical in shape. A probabilistic model for the shape matrices can be framed as follows:

$$S \sim F(\Sigma_k). \tag{1.5}$$

Here F is a probability distribution over the space of 2×2 symmetric positive definite matrices and Σ_k is the 'population average' shape matrix of the kth experimental condition. In these terms,

the hypothesis of no change in shape across G experimental conditions can be parametrized as

$$H_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_G. \tag{1.6}$$

The shape matrix contains information relating to a number of different aspects of the ellipse. Among these are (a) its determinant, d, which is proportional to the area of the ellipse; (b) the ratio of its maximum eigenvalue to its minimum eigenvalue, e, also known as the elongation or eccentricity of the ellipse, which is indicative of its shape; (c) the direction of the eigenvector corresponding to the maximum eigenvalue (major axis), θ, which is indicative of the orientation of the ellipse. It is possible to parametrize the shape matrix in terms of these quantities as

$$\Sigma = \sqrt{d} R(\theta) \begin{pmatrix} \sqrt{e} & 0 \\ 0 & 1/\sqrt{e} \end{pmatrix} R^{\mathrm{T}}(\theta), \tag{1.7}$$

where

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{1.8}$$

is Given's rotation matrix. A descriptive comparison of the data in terms of these statistics is shown in Fig. 2. From this figure, we can see that there is substantial inter-image variation in size and orientation distributions, but there do not appear to be any systematic differences across experimental conditions. However, for the elongation statistic, there does appear to be some increase in elongation in T4 relative to control and T1 conditions.

The neuroblastoma cells shown in Fig. 1 are grown in culture and are freely mobile in a fluid medium. Therefore the orientation of the nuclei or the cells may be arbitrary. In such cases, one would like to formulate a hypothesis which only focuses on the size and elongation aspects of the shapes. Additionally, statistics used for comparisons across experimental conditions should be invariant with respect to orientation. Clearly, the null hypothesis in (1.6) would not be appropriate for such a scenario, because it also implies equality in orientation. For other cell types, such as epithelial or neuronal tissue, orientation is an important consideration [12,5]. Even in these situations, however, a null hypothesis of the type (1.6) may not be appropriate. Due to possibly unknown changes in the relative orientations of the images during the imaging process, there may be no common orientation for the entire data set. In such situations, we would like the test statistic to be equivariant with respect to orientation of cells within each image, but invariant under rotation of the entire image. This could be termed as a *semi*-orientation-invariant comparison.

One possible approach to orientation-invariant comparisons is to analyse the size (d) and shape (e) statistics obtained from the decomposition given in (1.7). Such an approach is followed in [14]. The attractiveness of this approach lies in the fact that the analysis can be done using a variety of well-known methods for analysing univariate data. However, the difficulty with this univariate approach is that the measurements obtained by fitting the ellipse are not identically distributed. For example, the accuracies of the estimated elongation and orientation both depend upon the size of the ellipse, which behaves as a kind of sample size in this regard. In addition, the accuracy of the estimated orientation also depends on the elongation of the ellipse: it is easier to determine the orientation of an elongated object than that of a circular one. These difficulties were addressed in [14] by using weighted linear models, with the area of the ellipse as the weight. However this approach is somewhat *ad hoc*. Moreover, there is no straightforward extension of this method to handling the semi-invariant problem.

A more systematic approach to the invariance problem is by means of a modelling approach using nuisance parameters. In the case of rotation-invariant comparisons, our model for the shape

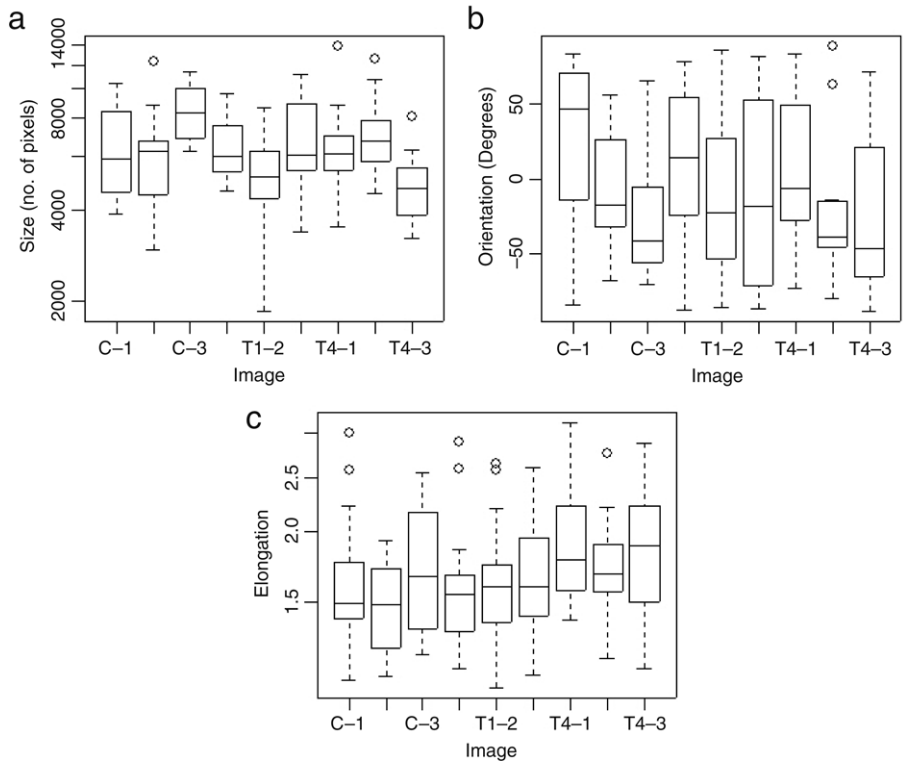

Fig. 2. Comparative box-plots of features of objects shown in Fig. 1 across images: (a) nucleus size, (b) nucleus orientation, (c) elongation.

matrices is

$$S \sim F\left(R(\theta)\Sigma_k R^{\mathrm{T}}(\theta)\right). \tag{1.9}$$

Here F and Σ_k are as in Eq. (1.5) and $R(\theta)$ is an arbitrary rotation applied to O. As our hypothesis test needs to be orientation invariant, we ignore the rotations in the statement of the null hypothesis and it stays the same as in (1.6). In the case of the semi-invariant comparison, a similar modelling framework can be applied. The only difference from (1.9) is that the arbitrary rotation is applied to each image, rather than each object:

$$S_{ijk} \sim F(R(\theta_{jk})\Sigma_k R^{\mathrm{T}}(\theta_{jk})). \tag{1.10}$$

The null hypothesis stays the same as before. In this paper, we will derive likelihood ratio tests (LRTs) for testing the null hypothesis (1.6) under each of the models (1.5), (1.9) and (1.10). In deriving these LRTs, we will treat the rotations as nuisance parameters, i.e. to be maximized over.

2. Likelihood ratio tests for shape matrices

In order to make explicit calculations of the LRT criterion, we need to make an assumption about the distribution of the shape matrices in the population. Our choice of distribution is guided

by the construction of the shape matrix S as defined in (1.4). It is identical to the construction of a sample variance–covariance matrix. Under the assumption that the points $p_i = (x_i, y_i)$ constitute a random sample from a bivariate normal distribution with covariance matrix Σ, the sample variance–covariance has a Wishart distribution [2, Th. 7.2.2] whose p.d.f. is given by

$$f\left(S_{ijk}\right) = \frac{1}{\sqrt{\pi}\, 2^{n_{ijk}} \prod_{l=1}^{n_{ijk}} \Gamma\left(\frac{n_{ijk}-l+1}{2}\right)} |\Sigma|^{-n_{ijk}/2} \left|n_{ijk} S_{ijk}\right|^{(n_{ijk}-3)/2} \times \exp\left\{\mathrm{tr}\left(-\frac{1}{2}\Sigma^{-1} n_{ijk} S_{ijk}\right)\right\}, \tag{2.1}$$

where $|\ |$ stands for determinant and tr stands for trace of a matrix. Γ is the gamma function.

For the analogy to be perfect, we require the points inside the object to constitute a random sample of a bivariate normal distribution. While this may be a tenuous assumption, the scatter of points that one would obtain from such a sample would indeed resemble an elliptical object. Therefore, we shall use the density in (2.1) for our LRT derivation. In this p.d.f., the parameter n_{ijk} is used as an indicator of the precision. For a random sample, the precision parameter is taken to be the size of the sample, which in this case would be the number of pixels in the object. However, object recognition algorithms would typically introduce spatial correlation amongst the chosen points. Therefore the effective sample size, and hence precision, is typically much lower. To reflect this fact, we have chosen our precision parameters as the square roots of the number of pixels in the object.

2.1. Absolute common orientation

The likelihood ratio test for testing the null hypothesis (1.6) against a general alternative based on the model (1.5) can be defined as

$$\mathrm{LRT} = \frac{\sup_{\omega} L\left(S_{ijk}, \Sigma_k\right)}{\sup_{\Omega} L\left(S_{ijk}, \Sigma_k\right)}, \tag{2.2}$$

where Ω is the entire parameter space and ω is the restricted parameter space under the null hypothesis. Under the null hypothesis, all experimental groups have a common population shape matrix, which we will call Σ. It is not unreasonable to assume that the object shape matrices are independent of each other, conditional on the fact that they all come from the same population. On the basis of this independence and the above specification of the p.d.f., we can construct product likelihoods for the numerator and denominator in Eq. (2.2). Up to a constant not involving the parameter Σ, the likelihood in the numerator of (2.2) can be written as

$$\begin{aligned} L\left(S_{ijk}, \Sigma\right) &= c \prod_{k=1}^{G} \prod_{j=1}^{q_k} \prod_{i=1}^{m_{jk}} |\Sigma_k|^{-n_{ijk}/2} \exp\left\{\mathrm{tr}\left(-\frac{1}{2}\Sigma^{-1} n_{ijk} S_{ijk}\right)\right\} \\ &= c\, |\Sigma|^{-n_{\cdots}/2} \exp\left\{\mathrm{tr}\left(-\frac{1}{2} n_{\cdots} \Sigma^{-1} \bar{S}_{\cdots}\right)\right\}, \end{aligned} \tag{2.3}$$

where $n_{\cdots} = \sum_{i,j,k} n_{ijk}$, $\bar{S}_{\cdots} = \frac{1}{n_{\cdots}} \sum_{i,j,k} n_{\mathrm{ijk}} S_{ijk}$ and c is a constant not involving Σ. The maximizer of (2.3) as a function of Σ is well-known in multivariate analysis. It occurs when Σ is

equal to $\bar{S}_{...}$ [Anderson (1980), Th. 3.2.1]. Therefore the value of the numerator in Eq. (2.2) can be written as

$$\sup_{\omega} L\left(S_{ijk}, \Sigma\right) = c\left|\bar{S}_{...}\right|^{-n_{...}/2} \exp\{-n_{...}\}. \tag{2.4}$$

Similarly, we can write the likelihood in the denominator of Eq. (2.2) as

$$\begin{aligned} L\left(S_{ijk}, \Sigma_j\right) &= c \prod_{k=1}^{G} \prod_{j=1}^{q_k} \prod_{i=1}^{m_{jk}} |\Sigma_k|^{-n_{ijk}/2} \exp\left\{\operatorname{tr}\left(-\frac{1}{2}\Sigma_k^{-1} n_{ijk} S_{ijk}\right)\right\} \\ &= c \prod_{k=1}^{G} |\Sigma_k|^{-n_{ijk}/2} \exp\left\{\operatorname{tr}\left(-\frac{1}{2} n_{..k} \Sigma_k^{-1} \bar{S}_{..k}\right)\right\}, \end{aligned} \tag{2.5}$$

where

$$n_{..k} = \sum_{i,j} n_{ijk} \quad \text{and} \quad \bar{S}_{..k} = \frac{1}{n_{..k}} \sum_{i,j} n_{ijk} S_{ijk}.$$

As in (2.3), the maximizer of the likelihood in (2.5) occurs when Σ_k equal to $\bar{S}_{.j}$. Therefore the value of the denominator in Eq. (2.2) can be written as

$$\sup_{\Omega} L\left(S_{ijk}, \Sigma_k\right) = c \exp\{-n_{..k}\} \prod_{k=1}^{G} \left|\bar{S}_{..k}\right|^{-n_{..k}/2}. \tag{2.6}$$

Combining Eqs. (2.2), (2.4) and (2.6), we get the LRT test statistic:

$$\text{LRT (SHAPE)} = \frac{c \exp\{-n_{..k}\} \prod_{k=1}^{G} \left|\bar{S}_{..k}\right|^{-n_{..k}/2}}{c\left|\bar{S}_{...}\right|^{-n_{...}/2} \exp\{-n_{...}\}} = \frac{\prod_{k=1}^{G} \left|\bar{S}_{..k}\right|^{-n_{..k}/2}}{\left|\bar{S}_{...}\right|^{-n_{...}/2}}. \tag{2.7}$$

In the special case that $n_{jk} = I_k = 1$ (one object per experimental condition), the test statistic derived in Eq. (2.7) reduces to the test statistic for testing equality of several covariance matrices under the multivariate normal model [Anderson, 1980, Section 10.2]. The null hypothesis will be rejected for large values of the LRT statistic.

Alternatively, we can break down the problem of comparing G experimental conditions to sets of pairwise comparisons. This is particularly meaningful if $G = 2$ or 3. In this case, the problem becomes analogous to that of comparing the covariance structure of two groups. For a pairwise comparison of the shape matrices of two populations, Roy [13] has proposed the test statistics to be the maximum and minimum eigenvalues of the matrix $S_1 S_2^{-1}$, where S_1 and S_2 are the respective sample covariance matrices. Here we are looking for values of the maximum and/or minimum eigenvalues which are far from 1. To extend this approach to the present problem, we need to use the test statistic $\bar{S}_{..K} \bar{S}_{..k}^{-1}$. However this approach will involve multiple comparisons when $G > 2$, so adjustment of the observed significance level will be necessary.

2.2. *Images with unknown rotation*

In this section, we assume that each image has been rotated by an unknown angle θ_{jk}, leading to the model (1.10). In this case, the LRT criterion can be written as

$$\mathrm{LRT}^r = \frac{\sup\limits_{\omega} L^r\left(S_{ijk}, \theta_{jk}, \Sigma\right)}{\sup\limits_{\Omega} L^r\left(S_{ijk}, \theta_{jk}, \Sigma_k\right)}. \tag{2.8}$$

The numerator of Eq. (2.8) can be written as

$$\begin{aligned} L^r\left(S_{ijk}, \theta_{jk}, \Sigma\right) &= c \prod_{i,j,k} |\Sigma|^{-n_{ijk}/2} \exp\left\{\mathrm{tr}\left(-\frac{1}{2} R^{\mathrm{T}}\left(\theta_{jk}\right) \Sigma^{-1} R\left(\theta_{jk}\right) n_{ijk} S_{ijk}\right)\right\} \\ &= c\,|\Sigma|^{-n_{\cdots}/2} \exp\left\{\mathrm{tr}\left(-\frac{1}{2} n_{\cdots} \Sigma^{-1} \bar{S}^r_{\cdots}\right)\right\}, \end{aligned} \tag{2.9}$$

where

$$\bar{S}^r_{\cdots} = \frac{1}{n_{\cdots}} \sum_{i,j,k} n_{ijk} R\left(\theta_{jk}\right) S_{ijk} R^{\mathrm{T}}\left(\theta_{jk}\right) = \frac{1}{n_{\cdots}} \sum_{j,k} n_{\cdot jk} R\left(\theta_{jk}\right) \bar{S}_{\cdot jk} R^{\mathrm{T}}\left(\theta_{jk}\right)$$

and

$$\bar{S}_{\cdot jk} = \frac{1}{n_{\cdot jk}} \sum_i S_{ijk}.$$

Maximization must now be done over the parameters Σ and θ_{jk}. Fortunately, the maximization with respect to each parameter can be done successively, i.e.,

$$\sup_{\theta_{jk}, \Sigma} L^r\left(S_{ijk}, \theta_{jk}, \Sigma\right) = \sup_{\theta_{jk}} \sup_{\Sigma} L^r\left(S_{ijk}, \theta_{jk}, \Sigma\right). \tag{2.10}$$

The maximization with respect to Σ can be done as previously:

$$\sup_{\Sigma} L^r\left(S_{ijk}, \Sigma\right) = c\left|\bar{S}^r_{\cdots}\right|^{-n_{\cdots}/2} \exp\{-n_{\cdots}\}. \tag{2.11}$$

Maximizing the quantity in Eq. (2.11) with respect to the θ_{jk} gives

$$\begin{aligned} \sup_{\theta_{jk}, \Sigma} L^r\left(S_{ijk}, \theta_{jk}, \Sigma\right) &= \sup_{\theta_{jk}} c\left|\bar{S}^r_{\cdots}\right|^{-n_{\cdots}/2} \exp\{-n_{\cdots}\} \\ &= c \exp\{-n_{\cdots}\} \left(\sum_{j,k} \min \sigma\left(\bar{S}_{\cdot jk}\right)\right)^{-n_{\cdots}/2} \left(\sum_{j,k} \max \sigma\left(\bar{S}_{\cdot jk}\right)\right)^{-n_{\cdots}/2}. \end{aligned} \tag{2.12}$$

Here $\sigma(A)$ denotes the set of eigenvalues of the matrix A. In the case of two-dimensional images, $\sigma(A)$ has two elements. A proof of Eq. (2.12) is given in the Appendix. The denominator of Eq. (2.8) can be rewritten as

$$\begin{aligned} L^r\left(S_{ijk}, \theta_{jk}, \Sigma_k\right) &= c \prod_{k=1}^{G} \prod_{j=1}^{q_k} \prod_{i=1}^{m_{jk}} |\Sigma_k|^{-n_{ijk}/2} \exp\left\{\mathrm{tr}\left(-\frac{1}{2} R^{\mathrm{T}}\left(\theta_{jk}\right) \Sigma_k^{-1} R\left(\theta_{jk}\right) n_{ijk} S_{ijk}\right)\right\} \\ &= c \prod_{k=1}^{G} |\Sigma_k|^{-n_{\cdot\cdot k}/2} \exp\left\{\mathrm{tr}\left(-\frac{1}{2} n_{\cdot\cdot k} \Sigma_k^{-1} \bar{S}^r_{\cdot\cdot k}\right)\right\}, \end{aligned} \tag{2.13}$$

where

$$\bar{S}^r_{..k} = \frac{1}{n_{..k}} \sum_{i,j} n_{ijk} R\left(\theta_{jk}\right) S_{ijk} R^{\mathrm{T}}\left(\theta_{jk}\right) = \frac{1}{n_{..k}} \sum_{j} n_{.jk} R\left(\theta_{jk}\right) \bar{S}_{.jk} R^{\mathrm{T}}\left(\theta_{jk}\right).$$

As previously, the maximization can be done in two steps. In the first step, we maximize over the Σ_k, yielding a result similar to Eq. (2.6):

$$\sup_{\Sigma_k} L^r\left(S_{ijk}, \theta_{jk}, \Sigma_k\right) = c \exp\{-n_{...}\} \prod_{k=1}^{G} \left|\bar{S}^r_{..k}\right|^{-n_{..k}/2}. \tag{2.14}$$

Maximizing the quantity in Eq. (2.14) with respect to the θ_{jk} gives

$$\begin{aligned} \sup_{\theta_{jk}, \Sigma} L^r\left(S_{ijk}, \theta_{jk}, \Sigma\right) &= c \exp\{-n_{...}\} \sup_{\theta_{jk}} \prod_{k=1}^{G} \left|\bar{S}^r_{..k}\right|^{-n_{..k}/2} \\ &= c \exp\{-n_{...}\} \prod_{k=1}^{G} \left(\sum_j \min \sigma\left(\bar{S}_{.jk}\right)\right)^{-n_{..k}/2} \left(\sum_j \max \sigma\left(\bar{S}_{.jk}\right)\right)^{-n_{..k}/2}. \end{aligned} \tag{2.15}$$

On substituting the expressions obtained in Eqs. (2.12) and (2.15), the LRT test statistic in Eq. (2.8) becomes

$$\mathrm{LRT}^r = \frac{\prod_{k=1}^{G} \left(\sum_j \min \sigma\left(\bar{S}_{.jk}\right)\right)^{\frac{n_{..k}}{2}} \left(\sum_j \max \sigma\left(\bar{S}_{.jk}\right)\right)^{\frac{n_{..k}}{2}}}{\left(\sum_{j,k} \min \sigma\left(\bar{S}_{.jk}\right)\right)^{\frac{n_{...}}{2}} \left(\sum_{j,k} \max \sigma\left(\bar{S}_{.jk}\right)\right)^{\frac{n_{...}}{2}}}. \tag{2.16}$$

2.3. Orientation-invariant comparison

In this section, we look at the situation where orientation is not an important facet of the shape. We say that each object is rotated by an angle θ_{ijk} (with respect to some common reference orientation). Under model (1.9), the LRT can be written as

$$\mathrm{LRT}^l = \frac{\sup_{\omega} L^l\left(S_{ijk}, \theta_{ijk}, \Sigma\right)}{\sup_{\Omega} L^l\left(S_{ijk}, \theta_{ijk}, \Sigma_k\right)}. \tag{2.17}$$

Under H_0, the likelihood given in the numerator of Eq. (2.17) can be written as

$$\begin{aligned} L^l\left(S_{ijk}, \theta_{ijk}, \Sigma\right) &= c \prod_{i,j,k} |\Sigma|^{-n_{ijk}/2} \exp\left\{\mathrm{tr}\left(-\frac{1}{2} R^{\mathrm{T}}\left(\theta_{ijk}\right) \Sigma^{-1} R\left(\theta_{ijk}\right) n_{ijk} S_{\mathrm{ijk}}\right)\right\} \\ &= c\,|\Sigma|^{-n_{...}/2} \exp\left\{\mathrm{tr}\left(-\frac{1}{2} n_{...} \Sigma^{-1} \bar{S}^l_{...}\right)\right\}, \end{aligned} \tag{2.18}$$

where

$$\bar{S}^l_{...} = \frac{1}{n_{...}} \sum_{i,j,k} n_{\mathrm{ijk}} R\left(\theta_{ijk}\right) S_{ijk} R^{\mathrm{T}}\left(\theta_{ijk}\right).$$

As before, maximization of $L^l\left(S_{ijk}, \theta_{ijk}, \Sigma\right)$ can be done in two steps. First,

$$\sup_{\Sigma} L^l\left(S_{ijk}, \Sigma\right) = c\left|\bar{S}^l_{...}\right|^{-n_{...}/2} \exp\{-n_{...}\}. \tag{2.19}$$

Then,

$$\begin{aligned}\sup_{\theta_{ijk},\Sigma} L^l\left(S_{ijk}, \theta_{ijk}, \Sigma\right) &= \sup_{\theta_{ijk}} c\left|\bar{S}^l_{...}\right|^{-n_{...}/2} \exp\{-n_{...}\} \\ &= c\exp\{-n_{...}\}\left(\sum_{i,j,k}\min\sigma\left(S_{ijk}\right)\right)^{-n_{...}/2}\left(\sum_{i,j,k}\max\sigma\left(S_{ijk}\right)\right)^{-n_{...}/2}.\end{aligned} \tag{2.20}$$

The denominator of Eq. (2.17) can be written as

$$\begin{aligned}L^l\left(S_{ijk}, \theta_{ijk}, \Sigma_k\right) &= c\prod_{k=1}^{G}\prod_{j=1}^{q_k}\prod_{i=1}^{m_{jk}}|\Sigma_k|^{-n_{\mathrm{ijk}}/2} \\ &\quad\times\exp\left\{\mathrm{tr}\left(-\frac{1}{2}R^{\mathrm{T}}\left(\theta_{ijk}\right)\Sigma_k^{-1}R^{\mathrm{T}}\left(\theta_{ijk}\right)n_{ijk}S_{ijk}\right)\right\} \\ &= c\prod_{k=1}^{G}|\Sigma_k|^{-n_{..k}/2}\exp\left\{\mathrm{tr}\left(-\frac{1}{2}n_{..k}\Sigma_k^{-1}\bar{S}^l_{..k}\right)\right\},\end{aligned} \tag{2.21}$$

where

$$\bar{S}^l_{..k} = \frac{1}{n_{..k}}\sum_{i,j} n_{ijk}R\left(\theta_{ijk}\right)S_{ijk}R^{\mathrm{T}}\left(\theta_{ijk}\right)$$

The first step of maximization of (2.21) yields

$$\sup_{\Sigma_k} L^r\left(S_{ijk}, \theta_{\mathrm{ijk}}, \Sigma_k\right) = c\exp\{-n_{...}\}\prod_{k=1}^{G}\left|\bar{S}^l_{..k}\right|^{-n_{..k}/2}. \tag{2.22}$$

Maximizing the quantity in Eq. (2.22) with respect to the θ_{ijk} gives

$$\begin{aligned}\sup_{\theta_{ijk},\Sigma} L^l\left(S_{ijk}, \theta_{ijk}, \Sigma\right) &= c\exp\{-n_{...}\}\sup_{\theta_{ijk}}\prod_{k=1}^{G}\left|\bar{S}^l_{..k}\right|^{-n_{..k}/2} \\ &= c\exp\{-n_{...}\}\prod_{k=1}^{G}\left(\sum_{i,j}\min\sigma\left(S_{ijk}\right)\right)^{-n_{..k}/2}\left(\sum_{i,j}\max\sigma\left(S_{ijk}\right)\right)^{-n_{..k}/2}.\end{aligned} \tag{2.23}$$

On substituting the expressions obtained in Eqs. (2.20) and (2.23), the LRT test statistic in Eq. (2.8) becomes

$$\mathrm{LRT}^l = \frac{\prod_{k=1}^{G}\left(\sum_{i,j}\min\sigma\left(S_{ijk}\right)\right)^{n_{..k}/2}\left(\sum_{i,j}\max\sigma\left(S_{ijk}\right)\right)^{n_{..k}/2}}{\left(\sum_{i,j,k}\min\sigma\left(S_{ijk}\right)\right)^{\frac{n_{...}}{2}}\left(\sum_{i,j,k}\max\sigma\left(S_{ijk}\right)\right)^{\frac{n_{...}}{2}}}. \tag{2.24}$$

3. Comparison of LRT statistics

By compiling results from the previous section, we are now in a position to compare the forms of the LRT under varying degrees of orientation invariance. For ease of comparison, these are restated below in a comparable format:

Orientation invariance

$$2\log\left(\mathrm{LRT}^{I}\right) = \sum_{k=1}^{G} n_{..k}\left\{\log\left(\sum_{i,j}\min\sigma\left(S_{\mathrm{ijk}}\right)\right) + \log\left(\sum_{i,j}\max\sigma\left(S_{ijk}\right)\right)\right\} - n_{...}\left\{\log\left(\sum_{i,j,k}\min\sigma\left(S_{ijk}\right)\right) + \log\left(\sum_{i,j,k}\max\sigma\left(S_{ijk}\right)\right)\right\}. \tag{3.1}$$

Images rotated

$$2\log\left(\mathrm{LRT}^{r}\right) = \sum_{k=1}^{G} n_{..k}\left\{\log\left(\sum_{j}\min\sigma\left(\bar{S}_{.jk}\right)\right) + \log\left(\sum_{j}\max\sigma\left(\bar{S}_{.jk}\right)\right)\right\} - n_{...}\left\{\log\left(\sum_{j,k}\min\sigma\left(\bar{S}_{.jk}\right)\right) + \log\left(\sum_{j,k}\max\sigma\left(\bar{S}_{.jk}\right)\right)\right\}. \tag{3.2}$$

Common orientation

$$2\log\left(\mathrm{LRT}\right) = \sum_{k=1}^{G} n_{..k}\left\{\log\left(\min\sigma\left(\bar{S}_{..k}\right)\right) + \log\left(\max\sigma\left(\bar{S}_{..k}\right)\right)\right\} - n_{...}\left\{\log\left(\min\sigma\left(\bar{S}_{...}\right)\right) + \log\left(\max\sigma\left(\bar{S}_{...}\right)\right)\right\}. \tag{3.3}$$

Eqs. (3.1)–(3.3) show that the LRTs are very similar in form, in that they are all contrasts of log eigenvalues of various shape matrices to the respective average. Shape matrices are symmetric positive semi-definite matrices, which form a cone in four-dimensional space: convex combinations of positive semi-definite matrices are also positive semi-definite matrices. The use of average shape matrices to study populations of elliptical objects has previously been used in the field of strain analysis, where the strain tensor is estimated by averaging the shape matrices of a distribution of elliptical geological markers [16]. In the present case, the amount of averaging follows the degree of invariance required. In the situation where there is a common absolute orientation across all images, the contrast is between (log eigenvalues of) groupwise average shape matrices against the overall average shape matrix. In the case where images are arbitrarily rotated relative to each other, the contrast is between various linear combinations of (log eigenvalues of) average shape matrices within each image. Thus the amount of averaging in this case is somewhere in between the orientation-invariant and absolute orientation scenarios.

We have already noted that the test statistic for equality of shape matrices with common orientation reduces to Box's M test for the equality of covariance matrices [4] for the case $n_{jk} = I_k = 1$. In the case of orientation invariance, eigenvalues are computed for the shape matrix of each object. For the case of invariance under nonsingular linear transformations, [Anderson (1980), sec. 10.6.2] considers the hypothesis $H_0 : \Sigma_1 = \Sigma_2$ for $G = 2$, again in the case $n_{jk} = I_k = 1$. In this case, the test statistic is based on the maximal invariant, which is the

set of eigenvalues of the matrix $S_1^{-1}S_2$. We note that there is no general extension of this result when $G > 2$ for covariance matrices of any dimension p. However, in the case when $p = 2$, which is the case for images, an extension of the nature of (3.1) can be obtained when the set of transformations is restricted to rotations.

A desirable invariance property shared by all three test statistics is the fact that they are invariant to changes in a common scale parameter within an image. This means that test statistics derived from different experiments which acquire images at different levels of magnification are comparable, *ceteris paribus*. Within an experiment, however, we require all images to be of the same magnification.

3.1. Simulation experiment

This section illustrates the validation of the test statistics described above in two simple simulation experiments. For simplicity, we restrict ourselves to the situation when $G = 2$, $I_k \equiv 1, n_k \equiv 20$ and $n_{ijk} \equiv 500$, i.e., two groups, one image per group, 20 objects per image and each object is of size 500. In this simplified setting, it is not meaningful to consider the images-rotated test statistic, so we restrict to a comparison between the common orientation and orientation-invariant test statistics. First, we consider an experiment where the only difference is a change in average orientation. The model generating the data is as described in (1.5). The average shape matrices are

$$\Sigma_a = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma_b = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}. \tag{3.4}$$

From its definition, it is evident that the orientation-invariant test (3.1) is not suited for comparisons where the only likely change across experimental conditions is in the orientation of the objects. This is reflected in Fig. 3(a), where we see that the distribution of this test statistic under $H_0 : \Sigma_1 = \Sigma_2 = \Sigma_a$ and the alternative $H_0 : \Sigma_1 = \Sigma_a; \Sigma_2 = \Sigma_b$, are the same. On the other hand, in Fig. 3(b), we can see a clear differentiation between the distributions of the common orientation test statistic under the same two hypotheses. In Fig. 3(e), we can see that the power function of the common orientation statistic in equal to 1 for almost every significance level. By contrast, the power function for the orientation-invariant test is approximately the line $y = x$, because the distributions under the null and the alternative are nearly identical.

When data are generated according to the model described in (1.9), we now demonstrate a potential scenario where the opposite may be true. Under H_0, data for both groups are generated from a Wishart distribution with average matrix $\Sigma_1 = 2I$ (a circle of radius 2). Under H_1, one group has the same distribution as under H_0 while for the other group data are generated from a Wishart distribution with average matrix that is given by

$$\Sigma_2 = R(\theta) \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} R^{\mathrm{T}}(\theta). \tag{3.5}$$

Here the rotation parameter θ is chosen as a random sample from a uniform distribution between $-\pi/2$ to $\pi/2$. This is a special case of the model described in (1.10). The distribution of both test statistics under the null hypothesis is shown in Fig. 3. Here we see that for both test statistics, there are differences between the distributions under H_0 and H_1. However, the discrimination between the two hypotheses is much clearer for the orientation-invariant test as opposed to the common orientation test. In Fig. 3(f), we can see that the power function for the orientation-invariant test is greater than the power function for the common orientation statistic at

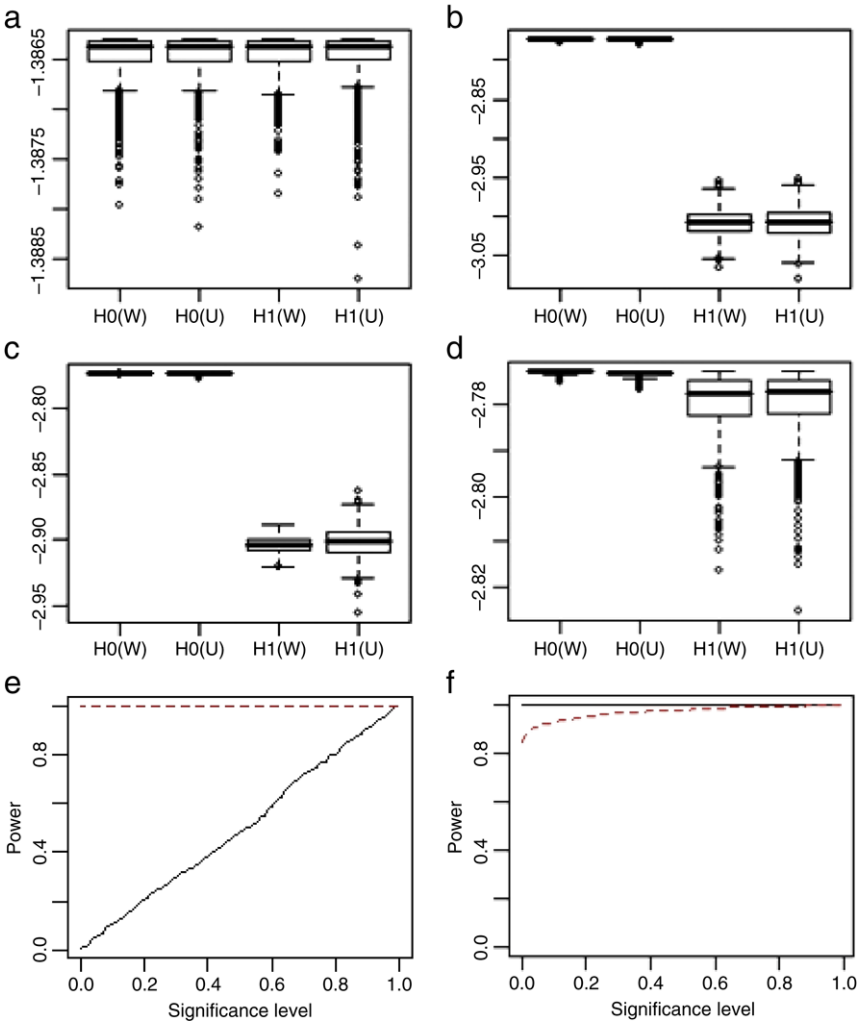

Fig. 3. Results of simulation experiment described in Section 3.1. Plots show the distribution of the test statistic under 1000 replications of the experiment for ((a), (c)) orientation-invariant LRT; ((b), (d)) common orientation LRT. In (a) and (b), the distribution is obtained under the null and alternative hypotheses given in Eq. (3.4). In (a) and (b), the alternative hypothesis is as given in Eq. (3.5), the null remains the same as in Eq. (3.4). In plots (a)–(d), the distribution of the test statistic is computed both under the Wishart (W) and Uniform (U) models. See the text for details. Plots (e) and (f) show the power as a function of significance level (α). The solid curve denotes the orientation-invariant test and the dashed curve denotes the common orientation test. Plot (e) corresponds to the distributions in plot (a) and (b) and plot (f) corresponds to the distributions in plot (c) and (d).

almost every significance level, although the difference diminishes at higher significance levels, which may not be of practical interest. This indicates that the orientation-invariant test would be more powerful in discriminating between the two alternatives. The reason why the invariant test does well is clear: there are differences in average axial ratio under H_0 and H_1. For the common orientation case, which relies of the average shape matrix, the situation is not so clear, because the average shape matrices under H_0 and H_1 are quite similar, modulo scale.

3.1.1. Robustness to choice of model

All the test statistics in this paper are derived under a Wishart model for the shape matrices of the objects. There is no scientific/probabilistic argument to support this assumption, although there is some empirical evidence of its validity for the data set at hand (see discussion). It is therefore important to study the efficacy of the tests proposed under alternate models. To this end, we consider a model where the lengths of major and minor axes are randomly sampled (independently of each other) from a uniform distribution with mean values given by diagonals of the mean shape matrix under the corresponding hypothesis, as in Eqs. (3.4) and (3.5). For the model in Eq. (3.5), the orientation parameter is chosen as before. This sampling model roughly corresponds to a measurement error model where axes are measured manually. A common value for the variance of the uniform distributions (i.e. the width of the interval) is chosen to approximately match the variability under the Wishart likelihood. From Fig. 3(a)–(d), we can see that the distributions of the test statistics under the Wishart and Uniform distributions are quite close under all the hypotheses under consideration.

4. Reference distribution of test statistics

In the special case $n_{jk} = I_k = 1$, certain results on the null distribution of LRTs similar to those derived in Section 3 are available. For instance, the distribution of the criterion (3.3) is shown to be a product of beta distributions in [Anderson (1980) Th. 10.4.2]. Asymptotic expansions of such criteria are also available [4]. For the invariant test statistic based on the eigenvalues of $S_1^{-1}S_2$, distributional results are reviewed in [6]. However, all of these results are derived under the Wishart assumption, which in the case of objects, may be a tenuous assumption. Moreover, they do not immediately extend to the invariant and semi-invariant tests described in this paper.

4.1. Unrestricted randomization

Improvements in computing power have popularized the construction of reference distributions of test statistics through randomization procedures. We describe one such method for the test statistics presented here. Let $S_1, \ldots, S_N$ ($N = \sum_k \sum_j n_{jk}$) denote the shape matrices of the objects in the study. Let $p(1), \ldots, p(N)$ denote a random permutation of the numbers $1, \ldots, N$. Then $S_{p(1)}, \ldots, S_{p(N)}$ constitutes a randomized redistribution of the data into the various groups. Using the randomization principle, the distribution of the test statistic under the null hypothesis (no difference between groups) can be obtained by repeatedly generating random permutations of the data and computing the test statistic. This method of generating the reference distribution works for the LRT with common orientation (3.3) because under the null hypothesis (1.6), the distribution of the shape matrices is exchangeable. In the case of orientation invariance the distribution of the shape matrices under the null hypothesis is $S_{ijk} \sim F\left(R\left(\theta_{ijk}\right) \Sigma R^{\mathrm{T}}\left(\theta_{ijk}\right)\right)$. The shape matrices are not identically distributed in this case, but the LRT (3.1) is based on just the eigenvalues of the shape matrices. As eigenvalues are invariant with respect to rotation, permutations of the shape matrices do not affect the distribution of the test statistic. The reference distributions of the test are shown in Fig. 4. From these distributions, we obtain an approximate significance level for the observed test statistic. For the common orientation test, the observed test statistic is -4005.709: this yields a p-value of 0.435. For the orientation-invariant test, the observed test statistic is -2095306, with a p-value is 0.028.

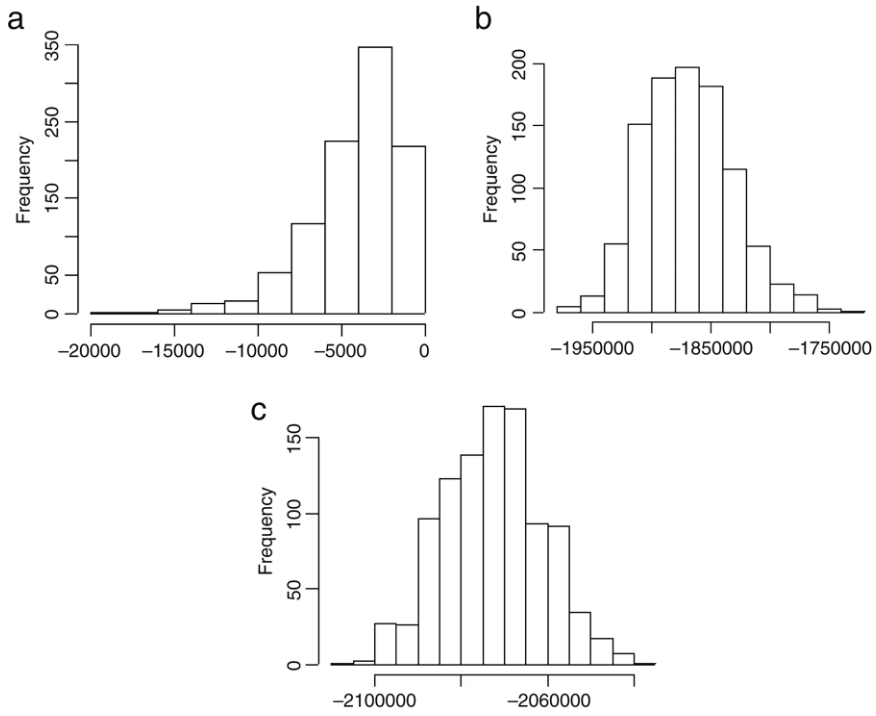

Fig. 4. Reference distributions of the LRTs under the null hypothesis, derived by randomization procedures described in Section 0; (a) orientation-invariant LRT, (b) common orientation LRT, (c) image rotation-invariant LRT.

4.2. Restricted randomization

In the case of rotated images, the distribution of matrices under the null hypothesis is $S_{ijk} \sim F\left(R\left(\theta_{jk}\right) \Sigma R^{\mathrm{T}}\left(\theta_{jk}\right)\right)$. The shape matrices are not identically distributed in this case, and unlike in the case of complete invariance, the test statistic (3.2) is not invariant with respect to rotations. This is because the test statistic is based on the average shape matrices $\bar{S}_{\cdot jk}$. If a permutation redistributes objects from image j to image j', the resulting objects in 'image' j' will have shape matrices from a mixture of distributions, which means that the null distribution of the permuted shape matrices will not be the same as the original arrangement. To avoid this problem, we need to restrict the randomization of objects so that objects are not redistributed across images. This effectively means that only entire images can be redistributed. As permutations within the same experimental condition do not change the value of the test statistic, the number of permissible distinct permutations can be quite small if the number of images is small. In the present example this number is ${}^{9}\mathrm{C}_{3}\,{}^{6}\mathrm{C}_{3} = 1680$, which is reasonable. The distribution of the test statistic for rotated images, given in (3.2), is obtained by this restricted randomization scheme and is shown in Fig. 3(c). From this reference distribution, the observed test statistic is -1842633, which yields a p-value of 0.769.

5. Discussion

We begin with a review of the scientific conclusions. The results of the hypothesis tests in Section 4 showed that there was a significant difference between the experimental conditions

under the orientation-invariant LRT. There were no significant differences across the groups according to the other two LRTs. This result is consistent with the findings of Roy Choudhury and Crotty [14], which indicated a significant difference in elongation across experimental conditions, but no significant differences in orientation. This result is indicative of the way the different LRTs in this paper should be interpreted: the orientation-invariant LRT is indicative only of changes of shape, whereas the other two LRTs are indicative of changes in both shape and orientation. However, these LRTs are designed for a situation where there is a single preferred orientation, such as in epithelial and neuronal cells. In many cell/tissue types, including the present example, there is no single preferred orientation. As indicated by the simulation example in Section 3.1, the orientation sensitive LRTs may fail to show significant differences under such circumstances. In this context, it would be of interest to exhaustively characterize scenarios under which such LRTs (or alternatives) would work well and others where they would not. Given such a characterization, it is possible that these tests could be used in conjunction to identify the scenario which gave rise to a particular set of significance results.

The choice of a Wishart model for shape matrices is partly motivated by convenience, because the likelihood can be maximized using analytic methods, as in much of classical multivariate distribution theory [2]. However, this choice can be partially empirically justified by looking at the data. It is difficult to visualize the distribution of the entire shape matrix as it resides in four-dimensional space, but one can look at marginal distributions: in particular, its eigenvalues. Under the Wishart assumption, the distribution of the eigenvalues of the shape matrix is asymptotically normal [2, Th. 13.5.1]. With $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2)$ denoting the diagonal matrix of eigenvalues of the mean shape matrix Σ, we have

$$\sqrt{n_{ijk}}\left(\sigma\left(S_{ijk}\right) - \Lambda\right) \to N_2\left(0, 2\Lambda^2\right). \tag{5.1}$$

Here $\sigma(S_{ijk})$ denotes the set of eigenvalues of the sample shape matrix. We note that the variance of this distribution is proportional to the square of its mean. By applying a variance stabilizing transform, in this case the log, we can obtain a homoscedastic distribution for the eigenvalues as follows [15, Th. 3.1 A]:

$$\sqrt{n_{ijk}}\left(\log\sigma\left(S_{ijk}\right) - \log\Lambda\right) \to N_2\left(0, c^2 I\right). \tag{5.2}$$

However, we note that the mean function for the eigenvalues depends on factors such as image and group, as in the alternate hypotheses in Section 3. We therefore correct for these factors by fitting a linear model to the log eigenvalues, as follows:

$$\log\sigma\left(S_{ijk}\right) = \mu + \alpha_k + \beta_{jk} + n_{ijk}^{-0.5}\varepsilon_{ijk}. \tag{5.3}$$

Here μ is the overall mean, α_k is the effect of the kth treatment, β_{jk} is the 'batch' effect of the jkth image, n_{ijk} is the size of the object and ε_{ijk} is i.i.d. $N(0, c^2 I)$ error. The distribution of the residuals from this model fitting process is shown in Fig. 5. We can see that both distributions look approximately Gaussian. This suggests that the assumption of a Wishart distribution for the shape matrices may indeed be tenable. We note, however, that this diagnostic method only captures a partial view of the distribution of the shape matrices.

The robustness exercise in Section 3.1.1 shows that the test statistics derived in this paper can be used in situations where the underlying model is not Wishart: almost similar distributions were obtained under a uniform model. A possible reason for this robustness lies in the form of the test statistics, which are essentially based on various linear combinations of the shape matrices

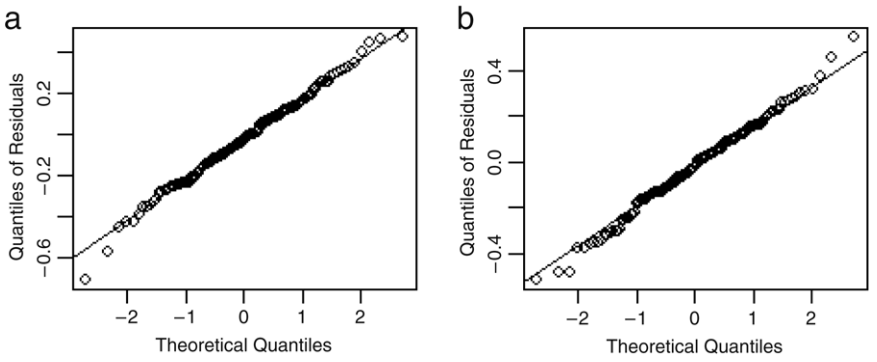

Fig. 5. Q–Q plots of the distribution of the residuals of the (a) major axes and (b) minor axes of the fitted ellipses from Fig. 1. The residuals are obtained by fitting the linear model (5.3) correcting for group and image specific effects.

and their eigenvalues. This suggests that the distribution of the test statistics might have some sort of limiting law with usual consistency properties. For instance, Shimamoto and Ikeda [16] have shown that, in the context of geologic strain estimation, the average shape matrix yields consistent estimates of the strain under a semi-parametric model. Another possible extension of this methodology would be to consider objects which are not elliptical in shape: it is possible that the present methodology may break down for such classes of objects.

A limitation of the analysis presented in this paper is that the measurement error model implied by the likelihood in (2.1) may not be realistic. This model suggests that the observed shape matrices arise as the sample variance covariance matrices of a random sample from a bivariate Gaussian distribution. In practice, the observed shape matrices are obtained from images using image segmentation routines such as edge detection or region growing [14], which behave unlike a random sampling scheme. Unfortunately, not much is known about the statistical properties of such algorithms, due to their highly adaptive nature. Our approach to addressing measurement error issues is somewhat *ad hoc*: we have used a weighting scheme where objects are weighted by square root of object size, as incorporated in the test statistics (3.1)–(3.3). In reality, the choice of weight will depend on the measurement process employed. To guard against misspecification of the precision parameter, a sensitivity analysis could be employed, whereby other possible candidates, such as log or other powers of object size are also considered. In the present example, the conclusions do not appear to be altered for the data considered when log object size is chosen as the precision parameter, although p-values are changed slightly.

Many current image cytometry data sets are acquired as 3D image stacks. For such data, one could try to extend the results to 3×3 shape matrices. Extending the results to such matrices is straightforward in most cases, except for the key result on minimizing the determinant given in the Appendix. The restricted randomization procedure indicated in Section 4.2 is only reasonable to use when there are sufficient numbers of images to permute. For instance, a minimum of 1900 permutations are required to obtain a coefficient of variation of less than 10% for the achieved significance level at a 5% level of significance ([9], Table 15.3). The required minimum number of permutations is inversely related to the observed p-value. In Section 4.2, the observed p-value is about 0.75, indicating that a much lower number of permutations would have sufficed. In situations where a smaller p-value is likely, but there are few images, an alternative re-sampling approach, such as a bootstrap, may be considered.

Acknowledgments

The authors would like to thank the anonymous referees for suggestions for improving the paper. The second author would like to thank Finbarr O'Sullivan for a helpful discussion of this problem.

Appendix. Minimization of a determinant

In this section we provide proofs of the statements made in Section 2 about the maxima of the expressions in Eqs. (2.11) and (2.14). The relevant quantities to be maximized are reciprocals of determinants of finite sums of rotational similarities of shape matrices. The problem, therefore, is to determine the minima of such determinants as the rotations are allowed to vary. Since a shape matrix is a real symmetric 2×2 matrix with non-negative eigenvalues, it is convenient to present the result in the following manner, bearing in mind that $\sigma(A)$ stands for the set of eigenvalues, i.e., the spectrum of a square matrix A.

Theorem 1. *Let $S_i, i = 0, 1, \ldots, n$, be $n+1$ non-negative-definite 2×2 matrices of real numbers. Denote by $\mathcal{R}$ the group of orthogonal 2×2 matrices of determinant one. Then*

$$\min\left\{\det\left(\sum_{i=0}^{n} R_i S_i R_i^t\right) : R_i \in \mathcal{R}, i = 0, 1, \ldots, n\right\} = \left(\sum_{i=0}^{n} \min \sigma(S_i)\right)\left(\sum_{i=0}^{n} \max \sigma(S_i)\right).$$

Proof. Since $R^t = R^{-1}$ and $\det R = 1$, $\forall R \in \mathcal{R}$, and $\det(AB) = \det A \det B$, for all square matrices A, B of the same size, it is clear that

$$\det\left(\sum_{i=0}^{n} R_i S_i R_i^t\right) = \det\left(S_0 + \sum_{i=1}^{n} R_0^t R_i S_i R_i^t R_0\right).$$

Also, since $\mathcal{R}$ coincides with the one-parameter multiplicative group of 2×2 rotation matrices of the form

$$R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

where $\theta \in \mathbb{R}$, so that

$$R(\theta + \phi) = R(\theta)R(\phi), \quad \theta, \phi \in \mathbb{R},$$

the problem reduces to determining the minimum value of the multi-variable function F defined on $\mathbb{R}^n$ by

$$F(\theta) = \det\left(S_0 + \sum_{i=1}^{n} R(\theta_i) S_i R(-\theta_i)\right),$$

where $\theta = (\theta_1, \theta_2, \ldots, \theta_n) \in \mathbb{R}^n$.

Since a sum of two non-negative-definite matrices is non-negative-definite, and a non-negative-definite matrix has a non-negative determinant, it is clear that F is non-negative and continuous. To determine its minimum value, we denote by a_i, b_i the eigenvalues of $S_i, i =$

$0, 1, \ldots, n$. Then $a_i, b_i \in [0, \infty), i = 0, 1, \ldots, n$. Moreover, we may, and do, assume that $a_i \geq b_i, i = 0, 1, \ldots, n$. Let

$$\Lambda_i = \begin{bmatrix} a_i & 0 \\ 0 & b_i \end{bmatrix}, \quad i = 0, 1, \ldots, n.$$

Since each S_i is orthogonally similar to a diagonal matrix, we may choose real numbers $\phi_i, i = 0, 1, \ldots, n$, so that

$$S_i = R(\phi_i)\Lambda_i R(-\phi_i), \quad i = 0, 1, \ldots, n.$$

Hence

$$\begin{aligned} F(\theta) &= \det\left(S_0 + \sum_{i=1}^{n} R(\theta_i) S_i R(-\theta_i)\right) \\ &= \det\left(R(\phi_0)\Lambda_0 R(-\phi_0) + \sum_{i=1}^{n} R(\theta_i)(R(\phi_i)\Lambda_i R(-\phi_i))R(-\theta_i)\right) \\ &= \det\left(R(\phi_0)\Lambda_0 R(-\phi_0) + \sum_{i=1}^{n} R(\theta_i + \phi_i)\Lambda_i R(-(\phi_i + \theta_i))\right) \\ &= \det\left(R(\phi_0)\left(\Lambda_0 + \sum_{i=1}^{n} R(\theta_i + \phi_i - \phi_0)\Lambda_i R(-(\phi_i + \theta_i - \phi_0))\right) R(-\phi_0)\right) \\ &= \det\left(\Lambda_0 + \sum_{i=1}^{n} R(\theta_i + \phi_i - \phi_0)\Lambda_i R(-(\phi_i + \theta_i - \phi_0))\right). \end{aligned}$$

In other words, writing $\phi = (\phi_1, \phi_2, \ldots, \phi_n)$, and identifying ψ with the constant vector $(\phi_0, \phi_0, \ldots, \phi_0)$, we see that

$$F(\theta + \psi - \phi) = \det\left(\Lambda_0 + \sum_{i=1}^{n} R(\theta_i)\Lambda_i R(-\theta_i)\right).$$

To proceed further, note that

$$\begin{aligned} R(\theta_i)\Lambda_i R(-\theta_i) &= \begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{bmatrix} \begin{bmatrix} a_i & 0 \\ 0 & b_i \end{bmatrix} \begin{bmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{bmatrix} \\ &= \begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{bmatrix} \begin{bmatrix} a_i\cos\theta_i & a_i\sin\theta_i \\ -b_i\sin\theta_i & b_i\cos\theta_i \end{bmatrix} \\ &= \begin{bmatrix} a_i\cos^2\theta_i + b_i\sin^2\theta_i & (a_i - b_i)\sin\theta_i\cos\theta_i \\ (a_i - b_i)\sin\theta_i\cos\theta_i & a_i\sin^2\theta_i + b_i\cos^2\theta_i \end{bmatrix}, \end{aligned}$$

for $i = 0, 1, \ldots, n$.

For a given vector $w \in \mathbb{R}^n$, define the functions f_w, g_w, h_w on $\mathbb{R}^n$ by

$$f_w(\theta) = \sum_{i=1}^{n} w_i \cos^2\theta_i, \; g_w(\theta) = \sum_{i=1}^{n} w_i \sin^2\theta_i, \; h_w(\theta) = \sum_{i=1}^{n} w_i \cos\theta_i \sin\theta_i,$$

so that

$$f_w(\theta) + g_w(\theta) = \sum_{i=1}^{n} w_i(\cos^2\theta_i + \sin^2\theta_i) = \sum_{i=1}^{n} w_i, \quad \forall\theta \in \mathbb{R}^n.$$

In this notation, we now have

$$\sum_{i=1}^{n} R(\theta_i)\Lambda_i R(-\theta_i) = \sum_{i=1}^{n} \begin{bmatrix} a_i \cos^2\theta_i + b_i \sin^2\theta_i & (a_i - b_i)\sin\theta_i\cos\theta_i \\ (a_i - b_i)\sin\theta_i\cos\theta_i & a_i \sin^2\theta_i + b_i\cos^2\theta_i \end{bmatrix}$$
$$= \begin{bmatrix} f_a(\theta) + g_b(\theta) & h_{a-b}(\theta) \\ h_{a-b}(\theta) & g_a(\theta) + f_b(\theta) \end{bmatrix},$$

where $a = (a_1, a_2, \ldots, a_n)$, $b = (b_1, b_2, \ldots, b_n)$. Hence

$$\begin{aligned} F(\theta + \psi - \phi) &= \det\left(\Lambda_0 + \sum_{i=1}^{n} R(\theta_i)\Lambda_i R(-\theta_i)\right) \\ &= \det\left(\begin{bmatrix} a_0 + f_a(\theta) + g_b(\theta) & h_{a-b}(\theta) \\ h_{a-b}(\theta) & b_0 + g_a(\theta) + f_b(\theta) \end{bmatrix}\right) \\ &= (a_0 + f_a(\theta) + g_b(\theta))(b_0 + g_a(\theta) + f_b(\theta)) - h_{a-b}^2(\theta) \\ &= a_0 b_0 + (a_0 f_b(\theta) + b_0 f_a(\theta)) + (a_0 g_a(\theta) + b_0 g_b(\theta)) \\ &\quad + (f_a(\theta) + g_b(\theta))(g_a(\theta) + f_b(\theta)) - h_{a-b}^2(\theta) \\ &= a_0 b_0 + (a_0 f_b(\theta) + b_0 f_a(\theta)) + (a_0 g_b(\theta) + b_0 g_b(\theta)) \\ &\quad + (f_a(\theta) + g_a(\theta))(g_b(\theta) + f_b(\theta)) \\ &\quad + (f_a(\theta) - f_b(\theta))(g_a(\theta) - g_b(\theta)) - h_{a-b}^2(\theta) \\ &= a_0 b_0 + (a_0 f_b(\theta) + b_0 f_a(\theta)) + (a_0 g_b(\theta) + b_0 g_b(\theta)) + \sum_{i=1}^{n} a_i \sum_{i=1}^{n} b_i \\ &\quad + (f_a(\theta) - f_b(\theta))(g_a(\theta) - g_b(\theta)) - h_{a-b}^2(\theta) \end{aligned}$$

Now, since $a_i - b_i \geq 0$, $i = 1, 2, \ldots, n$, we see that, for all $\theta \in \mathbb{R}$,

$$\begin{aligned} (f_a(\theta) - f_b(\theta))\,(g_a(\theta) - g_b(\theta)) - h_{a-b}^2(\theta) &= \sum_{i=1}^{n}(a_i - b_i)\cos^2\theta_i \sum_{i=1}^{n}(a_i - b_i)\sin^2\theta_i \\ &\quad - \left(\sum_{i=1}^{n}(a_i - b_i)\sin\theta_i\cos\theta_i\right)^2 \\ &\geq 0, \end{aligned}$$

by the Cauchy–Schwarz inequality. Furthermore, there is equality here if θ is the zero vector or the constant vector $(\pi/2, \pi/2, \ldots, \pi/2)$, modulo 2π. From this we infer that

$$\begin{aligned} F(\theta + \psi - \phi) &\geq a_0 b_0 + (a_0 f_b(\theta) + b_0 f_a(\theta)) + (a_0 g_b(\theta) + b_0 g_b(\theta)) + \sum_{i=1}^{n} a_i \sum_{i=1}^{n} b_i \\ &= a_0 b_0 + \sum_{i=1}^{n} a_i \sum_{i=1}^{n} b_i + \sum_{i=1}^{n}(a_0 b_i + b_0 a_i)\cos^2\theta_i \\ &\quad + \sum_{i=1}^{n}(a_0 a_i + b_0 b_i)\sin^2\theta_i \\ &= a_0 b_0 + \sum_{i=1}^{n} a_i \sum_{i=1}^{n} b_i + \sum_{i=1}^{n}(a_0 b_i + b_0 a_i) + (a_0 - b_0)\sum_{i=1}^{n}(a_i - b_i)\sin^2\theta_i \end{aligned}$$

$$\geq a_0 b_0 + \sum_{i=1}^{n} a_i \sum_{i=1}^{n} b_i + \sum_{i=1}^{n} (a_0 b_i + b_0 a_i)$$
$$= \left(a_0 + \sum_{i=1}^{n} a_i \right) \left(b_0 + \sum_{i=1}^{n} b_i \right),$$

with equality if θ is the zero vector.

It follows that

$$\begin{aligned}
\min F(\theta) &= F(\psi - \phi) \\
&= \det \left(\Lambda_0 + \sum_{k=1}^{n} \Lambda_k \right) \\
&= \det \begin{bmatrix} \sum_{i=0}^{n} a_i & 0 \\ 0 & \sum_{i=0}^{n} b_i \end{bmatrix} \\
&= \sum_{i=0}^{n} a_i \sum_{i=0}^{n} b_i \\
&= \sum_{i=0}^{n} \max \sigma(S_i) \sum_{i=0}^{n} \min \sigma(S_i).
\end{aligned}$$

This ends the proof.

References

[1] Y. Amit, U. Grenander, A. Piccioni, Structural image restoration through deformable templates, J. Amer. Statist. Assoc. 86 (1991) 376–387.
[2] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, 2nd ed., Wiley, 1984.
[3] F. Bookstein, Morphometric Tools for Landmark Data: Geometry and Biology, Cambridge University Press, 1991.
[4] G.E.P. Box, A general distribution theory for a class of likelihood criteria, Biometrika 36 (1949) 317–346.
[5] A. Chenn, S.K. McConnell, Cleavage orientation and the asymmetric inheritance of Notch1 immunoreactivity in mammalian neurogenesis, Cell 82 (1995) 631–641.
[6] S. Chu, K. Pillai, Power comparison of two-sided tests of equality of two covariance matrices based on six criteria, Ann. Inst. Stat. Math. 31 (1979) 182–205.
[7] T.R. Crimmins, A complete set of Fourier descriptors for two-dimensional shapes, IEEE Trans. Syst. Man Cybern. 12 (1982) 848–855.
[8] S. Crotty, Morphological correlates of cell death in a model of Parkinson's disease, M.Sc. Thesis, Department of Anatomy, UCC, 2005.
[9] B. Efron, R. Tibshirani, An Introduction to the Bootstrap, Chapman and Hall, 1993.
[10] C. Glasbey, G. Horgan, Image Analysis for the Biological Sciences, Wiley, 1995.
[11] T. Pavlidis, Polygonal approximations by Newton's method, IEEE Trans. Comput. 26 (1977) 800–807.
[12] F.M. Peralta, T.B. Casale, Orientation and presence of epithelium are key to endotoxin-induced neutrophil migration, Eur. Respir. J. 11 (1998) 1053–1059.
[13] S.N. Roy, On a heuristic method of test construction and its use in multivariate analysis, Ann. Math. Statist. 24 (1953) 220–238.
[14] K. Roy Choudhury, S. Crotty, Semi automated image cytometry, in: Proceedings of Integrating Health Research – Sharing Visions, UCC, Cork, Nov. 2005, 2005.
[15] R. Serfling, Approximation Theorems in Mathematical Statistics, Wiley, 1980.
[16] T. Shimamoto, Y. Ikeda, A simple algebraic method for strain estimation from deformed ellipsoidal objects. 1. Basic theory, Tectonophysics 36 (1976) 315–337.